Incorporation of Unsaturated Isoleucine Analogues into Proteins *In Vivo*

Thierry Michon,[a] Francis Barbot[b] and David Tirrell[c]

[a]UNITE DE BIOCHIMIE ET DE TECHNOLOGIE DES PROTEINES, RUE DE LA GERAUDIERE, INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE, 44316 NANTES CEDEX 03, FRANCE
[b]UNIVERSITE DE POITIERS, BAT. GON AV. DU RECTEUR PINEAU, 86022 POITIERS CEDEX, FRANCE
[c]DIVISION OF CHEMISTRY AND CHEMICAL ENGINEERING, 210–41 CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA 91125, USA

1 Abstract

The translational activity of various unsaturated analogs of L-isoleucine was evaluated using an *Escherichia coli* strain auxotrophic for isoleucine. It was observed that the alkene [2-amino-3-methyl-4-pentenoic acid (**2**)] and alkyne [2-amino-3-methyl-4-pentynoic acid (**3**)] derivatives of L-isoleucine can support protein synthesis at levels approximately 50% of that observed in cultures supplemented with isoleucine. However, no incorporation of the αC or βC methylated derivatives could be detected. In order to examine the stereoselectivity of incorporation, the (2*S*, 3*S*) and (2*S*, 3*R*) diastereomers of **2** and **3** were prepared. The extents of isoleucine substitution *in vivo* were 80% and 70% for (2*S*, 3*S*)-**2** and (2*S*, 3*S*)-**3**, respectively, under the conditions examined in this study.

2 Introduction

The *in vivo* incorporation into proteins of amino acids analogs bearing non-biological chemical reactivity within their side chain would allow a completely new chemistry of proteins. For instance, this would have applications in the design of new materials by combining proteins to synthetic polymers, nucleic acids or carbon hydrates. In the past it was shown that *E. coli* is extremely permissive for the incorporation of artificial amino acids (see ref. 1 for review).

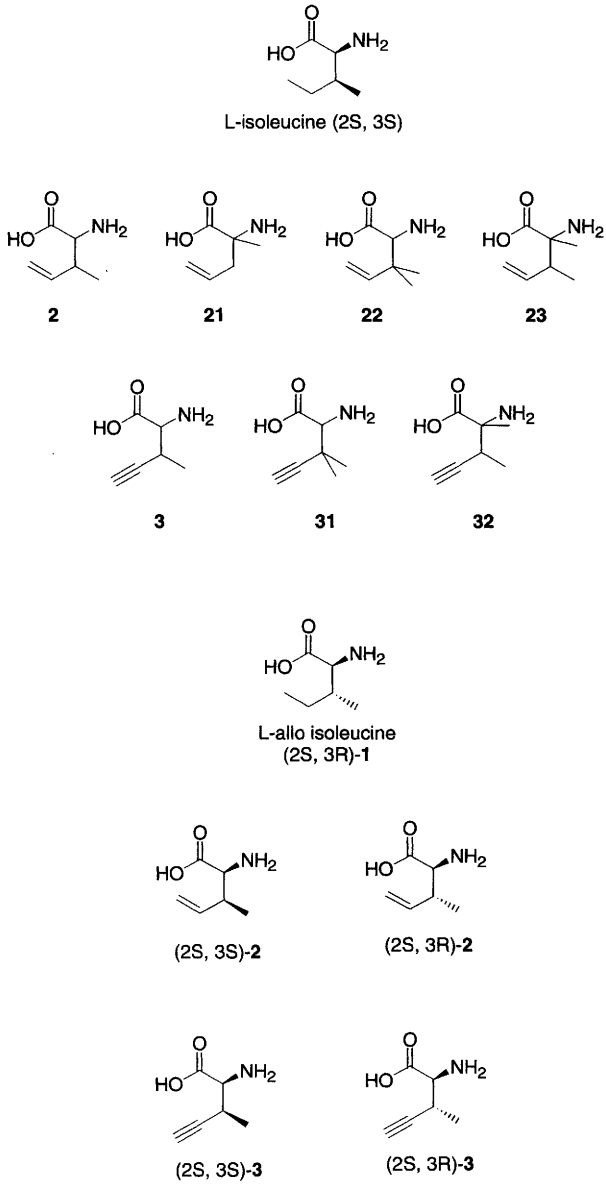

Figure 1 *Structures of the isoleucine analogs used in this study*

One of the reasons probably lies within the fact that in the course of evolution the selection pressure never applied to these amino acid analogs. As the laws governing protein structure and function emerge, it is becoming increasingly productive to design 'artificial proteins' as building blocks for new kinds of supra-molecular chemistry.[2,3] Optimizing such building blocks requires control of the driving forces that direct protein folding and assembly. This is achieved

largely by controlling the physical and chemical properties of the amino acid side chains of the protein of interest.[4] In an effort to increase the range of chemical function that can be incorporated into proteins, we have focused our investigations on the translational activity of unsaturated amino acid analogs because of the versatile chemistry of alkenes and alkynes.[5] For example, supramolecular structures made up of weakly hydrogen bonded cyclic peptides can be stabilized through inter-peptide cross-linking[6] utilizing ruthenium-catalyzed ring-closing metathesis of pendant alkene functions.[7] We have recently shown that homoallylglycine and homopropargylglycine can be incorporated *in vivo* into recombinant proteins.[8–10] In this paper we examine the incorporation of unsaturated isoleucine analogs (Figure 1) into a heterologous protein, mouse dihydrofolate reductase (mDHFR), over-expressed in an *Escherichia coli* host. Analogues **2** and **3** can be efficiently incorporated into *E. coli* proteins, and incorporation is stereospecific, favoring the (2*S*, 3*S*) diastereomers. The results are discussed from the standpoint of the insensitivity of the isoleucyl-tRNA synthetase (IleRS) editing mechanism to non-canonical amino acids.

3 Experimental Section

3.1 Amino Acid Synthesis

Diastereomers mixtures (Figure 1) of 2-amino-3-methyl-4-pentenoic acid (**2**), 2-amino-2-methyl-4-pentynoic acid (**21**), 2-amino-3-dimethyl-4-pentenoic acid (**22**), 2-amino-2,3-dimethyl-4-pentenoic acid (**23**), 2-amino-3-methyl-4-pentynoic acid (**3**), 2-amino-3-dimethyl-4-pentynoic acid (**31**), 2-amino-2,3-dimethyl-4-pentynoic acid (**32**) were prepared according to Aidene *et al.*[11] The regioselective reaction between α-unsaturated organozincs and *N*-(phenylsulfanyl)iminoesters was used as a starting step to obtain the α-aminoesters.[11]

Detailed procedures for the preparation of (**2**) and (**3**) pure diastereomers will be described elsewhere (Michon *et al.*, manuscript in preparation).

For all the compounds used, ^{1}H NMR spectra were in agreement with the expected structures.

3.2 Construction of an *E. coli* Expression Strain

An *E. coli* strain (AIV_s) that is auxotrophic for valine and isoleucine was prepared. The procedure will be described elsewhere (Michon *et al.* manuscript in preparation). An *in vivo* test system was designed in order to determine the extent of incorporation of isoleucine analogs into an over expressed reporter protein. Plasmid pQE15 (Qiagen) carries a gene encoding mouse dihydrofolate reductase (mDHFR) under control of a strong bacteriophage T5 promotor that is recognized by *E. coli* RNA polymerase. A repressor-binding site has been introduced downstream from the T5 promotor on pQE15 that allows a programmed induction of DHFR expression by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to the medium. The gene encoding mDHFR also encodes an N-terminal hexahistidine sequence to permit protein purification by

immobilized metal affinity chromatography. pQE15 confers ampiciline resistance. The auxotrophic *E. coli* strain AIV_s was transformed with pQE15 to give AIV_s/pQE15.[12] The repressor plasmid pLys-IQ modified from pLys-S (Novagen, Madison, WI, USA) to contain the $LacI^q$ gene encoding the lac repressor was then introduced into the AIV_s/pQE15 strain to give the AIV_s-IQ/pQE15 strain. This strain was made in an attempt to prevent 'leaky' expression of the DHFR gene before induction (see results). This strain was used for all the assays.

3.3 Protein Expression

To test for analog incorporation 5 mL of M9AA medium supplemented with ampicillin (200 μg/mL), chloramphenicol (35 μg/mL), 2 mM $MgCl_2$, 100 μM $CaCl_2$, 0.2% glucose and 0.5 mg thiamin chloride were inoculated with a single colony of AIV_s-IQ/pQE15. After overnight growth at 37°C the culture was diluted with fresh M9 medium in order to obtain an OD_{600} of 0.1. When the OD_{600} reached 1 (after about 3.5 hours) the cells were sedimented (5000 g, 10 min, 4°C) and washed twice with 0.9% NaCl. The cells were resuspended in 50 mL fresh M9 medium supplemented with the 19 amino acids (16 mg/L) but lacking isoleucine. Tests tubes containing 10 mL aliquots of this culture were prepared and supplemented with 250 μL water (negative control), 40 mg/L L-isoleucine (2*S*, 3*S*/2*R*, 3*R*) (positive control), or 80 mg/L of each of the analogs. After 10 min of growth DHFR expression was induced by addition of IPTG at a final concentration of 0.4 mM. The culture were grown at 37°C for 4 hours and 1 mL aliquots were spun down. Pellets were suspended in 50 μL of a 10 mM $mgCl^2$ solution containing 5 μg/mL DNAse and 10 μg/mL RNAse. The suspensions were frozen, thawed and sonicated prior to electrophoresis. In each case the remaining 9 mL were centrifuged and the pellets stored at −20°C overnight before DHFR purification. Cellular proteins were resolved by SDS PAGE, and mDHFR was detected by western blotting with antibodies raised against the histidine tag (Qiagen, Inc., Santa Clarita, CA, USA).[12]

3.4 DHFR Purification

Pellets were thawed for 30 min, and resuspended in 600 μL of buffer (6 M guanidine-HCl, 0.1 M NaH_2PO_4, 0.01 M Tris/HCl, pH 8). The mixture was vigorously shaken at room temperature for 1 h. The cell debris was sedimented (10 000 g, 30 min at room temperature) and the lysate was submitted to affinity chromatography on pre-packed Nickel columns according to the Ni-NTA spin column procedure described by Qiagen. The DHFR fractions were recovered in a buffer containing 8 M urea, 0.1 M NaH_2PO_4 and 0.01 M Tris/HCl pH 4.5. UV spectra of the samples were taken and the amount of DHFR obtained was quantified using $\varepsilon = 30500\ M^{-1}\ cm^{-1}$ at 280 nm.[13]

3.5 Estimation of the Level of Isoleucine Replacement into DHFR

Amino acid analysis were obtained from the Beckman Research Institute, Divi-

sion of Immunology, City of Hope, Duarte CA. The extent of replacement of isoleucine was estimated based on the diminution of isoleucine from its expected value. MALDI-TOF spectra were performed on a Voyager-DE STR spectrometer using 3,5-dimethoxy–4-hydroxy-cinnamoic acid as matrix (Mass Spectrometry Center of the Beckman Institute, California Institute of Technology).

4 Results

4.1 Analog Incorporation

Incorporation of selected amino acid analogs into proteins made *in vivo* has been known for many years.[1,14–16] In our study an *E. coli* strain auxotrophic for isoleucine was used to assay the extent of *in vivo* incorporation of isoleucine analogues into mouse dihydrofolate reductase (mDHFR), a test protein readily over expressed in bacterial cultures. It was expected that the *in vivo* incorporation of analogs into the proteins would kill the cells (cell lysis). In order to estimate the permissivity of *E. coli* protein biosynthesis, we fed the cells with diastereomer mixtures of isoleucine analogs. As shown in Figure 2, mDHFR was detected only when the culture medium was supplemented with **2** and **3**. It was shown earlier that amino acids with bulkier side chains than isoleucine could be accommodated in the binding site of the synthetase,[17,18] but none of the analogs bearing a double methylation on the βC carbon gave detectable amounts of mDHFR.

Among the 20 proteinogenic amino acids, only L-isoleucine and L-threonine carry a symmetry center in their side chain and only (2*S*, 3*S*)-isoleucine and (2*S*, 3*R*)-threonine are found in proteins. When L-isoleucine was replaced in the culture medium by L-alloisoleucine, the (2*S*, 3*R*)-**1** diastereomer of isoleucine, the cell growth slowed dramatically, without evidence of cell lysis. As western blotting did not show detectable DHFR expression in such cultures, we conclude that (2*S*, 3*R*)-**1** is incorporated into protein slowly if at all under our experimental conditions (Figure 3B, lane 5). However, small amounts of DHFR could be isolated from such culture (*ca.* 8% of the level of expression obtained in media supplemented with L-isoleucine, see Table 1). DHFR synthesis was not due to leaky expression before induction, as uninduced cells did not produce DHFR in media supplemented with L-isoleucine (Figure 3B, lane 3). Instead it appears that

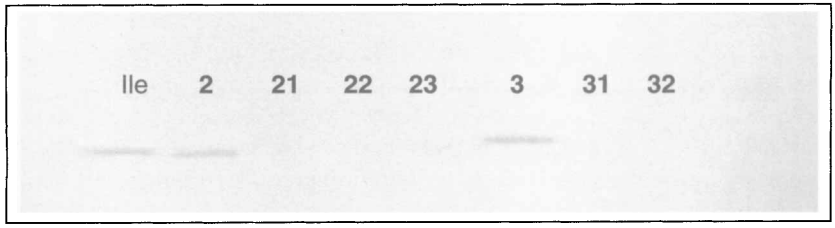

Figure 2 *Western blot obtained from a SDS PAGE of* E. coli *proteins 4 hours after medium shift. Amino acids added as indicated at the top of the figure. DHFR was detected with antibodies raised against the amino terminal hexahistidine tag*

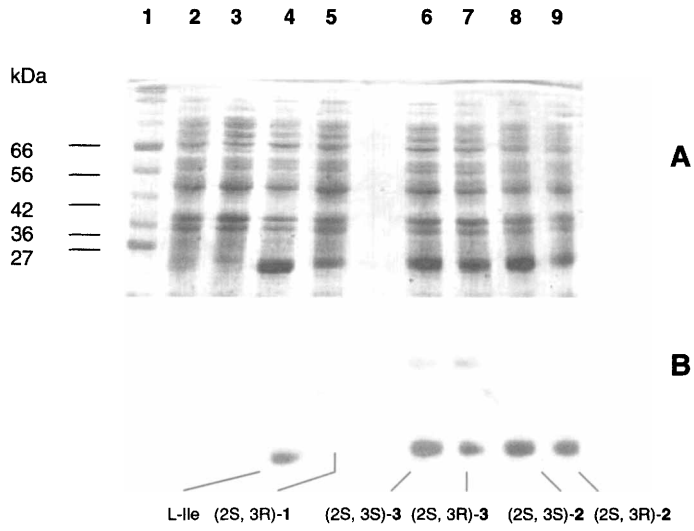

Figure 3 (*A*) *SDS PAGE of* E. coli *proteins 4 hours after medium shift. Lane 1: molecular weight standards; 2: no isoleucine added, induction; 3: isoleucine added, no induction. Lanes 4 to 9: analogs added as indicated at the bottom of the figure.* (*B*) *Autoradiogram of a western blot obtained from the gel above. DHFR was detected with antibodies raised against the amino terminal hexahistidine tag*

the host can provide a small pool of L-isoleucine, probably by means of intracellular proteolysis.

In order to compare the efficiency of the methyl position carried by βC on the incorporation, we prepared pure diastereomers of **2** and **3**.

(2*S*, 3*S*)-**2** did not support cell growth and was toxic to the host as shown by a drop in OD_{600} signaling cell lysis. Most importantly, (2*S*, 3*S*)-**2** supported synthesis of DHFR (Figure 3B, lane 8). Approximately 80% of the L-isoleucine in DHFR was replaced by the analog, as determined by amino acid analysis (Table 1). In contrast to (2*S*, 3*S*) the (2*S*, 3*R*) analog neither supported cell growth nor caused measurable lysis. DHFR could be isolated after induction of cultures supplemented with (2*S*, 3*R*)-**2** (Figure 3B lane 9), but in amounts too low to permit estimation of replacement efficiency (Table 1).

A fraction of **3** enriched with the (2*S*, 3*S*) isomer (diastereomeric ratio 8.6/1.4) was tested. This fraction did not support growth but instead caused cell lysis (Table 1), confirming the toxicity of the analog previously reported.[19] When DHFR was expressed in media supplemented with (2*S*, 3*S*)-**3** (Figure 2B lane 6) approximately 70% of the isoleucine was replaced by the analog (Table 1). When a fraction of **3** enriched in the (2*S*, 3*R*) diastereomer (diastereomeric ratio 8.3/1.7) was used the amount of DHFR recovered was too low to allow determination of the extent of incorporation.

MALDI-TOF mass spectra were recorded on the purified DHFR samples. The difference in mass obtained between native DHFR and its modified forms

Table 1 *Incorporation of isoleucine analogs into DHFR*

amino acid	*DHFR Yield*[a] μg/mL	$\overline{M_n}$[b]	% *replacement* AAA[c]	% *replacement* MALDI-TOF[d]
L-isoleucine (2*S*, 3*S*)	29.5	24044		
(2*S*, 3*R*)-**1**	2.4	ND	ND	ND
(2*S*, 3*S*)-**2**	16.4	24018	81	86
(2*S*, 3*R*)-**2**	1.35	ND	ND	ND
(2*S*, 3*S*)-**3**	11	24007	72	70
(2*S*, 3*R*)-**3**	ND	ND	ND	ND

[a]Yield of DHFR after purification from 10 mL culture, as determined by absorption at 280 nm.
[b]When mass spectra (MALDI-TOF) were recorded, all samples gave a main peak (85 to 95% of the signal intensity). The m/z value of this main peak was used to estimate $\overline{M_n}$, the average mass of DHFR.
[c]Extent of replacement of isoleucine as determined by amino acid analysis.
[d]Extent of replacement of isoleucine as estimated from the difference in $\overline{M_n}$ obtained between native DHFR and its modified forms.

were consistent with the % replacement determined by amino acid composition (data not shown).

5 Discussion

The first critical step in the incorporation of amino acids analogs is their uptake from the culture medium; the analog must be transported across the cytoplasmic membrane either by the machinery used for the uptake of its natural counterpart or by other import machinery. In the case of non polar isoleucine analogs this step might not be limiting as isoleucine and other non polar amino acids are likely to cross the phospholipids bilayer by simple diffusion. However, the leucine-isoleucine-valine (LIV)-binding proteins present at the surface of *E. coli* reveals great sequence similarity.[20] It was recently showed that the specificity of these systems might be weak enough to tolerate closely structurally related amino acids.[21]

In a second step the analog has to be coupled to a tRNA species by an aminoacyl-tRNA synthetase and must circumvent the editing pathways that normally limit misacylation of tRNAs. The selectivity (*s*) of an aminoacyl-tRNA synthetase toward an amino acid is defined as the ratio of the rate of editing to the rate of activation. It is noteworthy that the Met-tRNA synthetase exhibits a very high selectivity towards homocysteine ($s = 11000$) and norleucine ($s = 1000$) which are both biological amino acids present in the cell.[22] In contrast it was previously demonstrated in our group that the synthetase is able to mis-charge Met-tRNA with at least three unsaturated methionine analogues.[8,9] The editing mechanism of *E.coli* isoleucyl-tRNA synthetase (IleRS) has been extensively studied.[23,24] Its selectivity for natural amino acids is high, ranging from $s = 6000$ for valine to $s = 8.5 \times 10^6$ for alanine.[22] IleRS possesses two sites: one

for the binding of the amino acid prior to its activation through the formation of the AMP–AA phosphoester bond; the other, the editing site, for the hydrolysis of this bond when amino acids smaller than isoleucine (which easily fit into the binding pocket) have been inappropriately acylated.[25,26] Analogs **2** and **3** tested in this study appear to circumvent the editing mechanism of IleRS because they are too large to fit into the editing site. Our results show that IleRS is stereoselective as only (2*S*, 3*S*) isoleucine analogs are incorporated into protein at measurable rate. This is in agreement with previous binding studies which demonstrated that L-2 amino-3*S*-methylhexanoic acid binds to IleRS ($K_a = 20\ mM^{-1}$) with a stronger affinity than its diastereomer L-2 amino-3*R*-methylhexanoic acid ($K_a = 0.6\ mM^{-1}$).[17]

Finally, the analog may be edited at the ribosome level. If the editing mechanism of the aa-tRNA synthetase seems to insure a weak discrimination between natural and artificial amino acids one should expect a more efficient discrimination at the ribosome level. Studies performed *in vitro* show that this step is probably permissive enough.[27–30] The misscharged tRNA must avoid discrimination by elongation factor Tu (EFTu) (Figure 4). According to this scheme both the rate of GTP hydrolysis (k_2) and the rate of the EFTu.GDP complex dissociation with the ribosome (k_3) are defined as internal kinetic standard constants which do not depend on the presence of any cognate or non-cognate amino acid-tRNA. By contrast the dissociation constants between the ribosome and either the ternary complex EFTu.GTP.aa-tRNA (governed by k_{-1}) or the aa-tRNA after GTP conversion to GDP (governed by k_4) depend on the strength of the binding between the ribosome and the aa-tRNA (see ref. 31 for review). It is likely that the efficiency of transfer of the non-cognate amino acid to the on-growing peptidic chain is much lower than in the case of the natural amino acid. In the experiments reported here, the medium shift method prevents

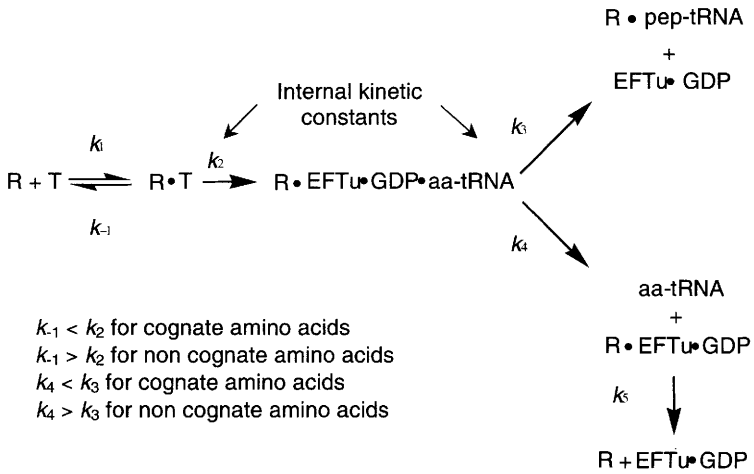

Figure 4 *Minimal mechanism by which aa-RNs are incorporated into nascent protein (see discussion for comments). AatRNA, aminoacyl tRNA; EFTu, elongation factor; GTP, guanosine triphosphate; GDP, guanosine diphosphate; R, ribosome; T, ternary complex (aa-tRNA•GTP•EFTu). Redrawn from Thompson* 1988[31]

competition by Ile and high level of incorporation of the analogs were obtained. However, it has been reported that furanomycin, an analogue of isoleucine, is readily incorporated into proteins.[18] Interestingly, in spite of large structural discrepancies between furanomycin and isoleucine, the equilibrium constant of the ternary complex formation (furanomycyl~tRNA•EFTu•GTP) equals that for isoleucyl~tRNA•EFTu•GTP formation. In the course of evolution nature selected editing processes on the basis of naturally occurring amino acids, which are present in the cell. This seems to open relatively large possibilities for the incorporation of artificial amino acids making use of the natural machinery of the cell.

Acknowledgment

The authors thank NATO for its financial support to T. M.'s year-long sabbatical. We are grateful to Kristi Kiick for fruitful discussions.

References

1. G. Hortin and I. Boime, *Methods in Enzymology*, 1983, **96**, 777.
2. M. Krejchi, E. Atkins, A. Waddon, M. Fournier, T. Mason and D. Tirrell, *Science*, 1994, **265**, 1427.
3. S. Sakamoto, I. Obataya, A. Ueno and H. Mihara, *J. Chem. Soc., Perkin Trans.*, 1999, **2**, 2059.
4. M. R. Ghadiri and C. Choi, *J. Am. Chem. Soc,*. 1990, **112**, 1630.
5. T. Michon and D. A. Tirrell, *Biofutur*, 2000, **197**, 34.
6. T. Clark and M. R. Ghadiri, *J. Am. Chem. Soc.*, 1995, **117**, 12364.
7. S. B. Nguyen, R. H. Grubbs and J. W. Ziller, *J. Am. Chem. Soc.*, 1993, **115**, 9858.
8. J. van Hest and D. Tirrell, *FEBS Lett.*, 1998, **428**, 68.
9. J. van Hest, K. Kiick and D. Tirrell, *J. Am. Chem. Soc.*, 2000, **122**, 1282.
10. K. L. Kiick, J. C. van Hest and D. A. Tirrell, *Angew. Chem. Int. Ed. Engl.*, 2000, **39**, 2148.
11. M. Aidene, F. Barbot and L. Miginiac, *J. Organometallic Chem.*, 1997, **534**, 117.
12. J. Sambrook, E. F. Fritsch and T. Maniatis, 'Molecular Cloning: A Laboratory Manual', 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13. J. Andrews, C. A. Fierke, B. Birdsall, G. Ostler, J. Feeney, G. C. Roberts and S. J. Benkovic, *Biochemistry*, 1989, **28**, 5743.
14. W. Hendrickson, J. Horton and D. Lemaster, *EMBO J.*, 1990, **9**, 1665.
15. E. D. Fenster and H. S. Anker, *Biochemistry*, 1969, **8**, 269.
16. T. W. Tuve, and H. H. Williams, *J. Am. Chem. Soc.*, 1957, **79**, 5830.
17. J. Flossdorf, H.-J. Pratorius and M.-R. Kula, *Eur. J. Biochem.*, 1976, **66**, 147.
18. T. Kohno, D. Kohda, M. Haruki, S. Yokoyama and T. Miyazawa, *J. Biol. Chem.*, 1990, **265**, 6931.
19. H. Gershon, J. Shapira, J. S. Meek, and K. Dittmer, *J. Am. Chem. Soc.*, 1954, **46**, 3484.
20. R. Tam and M. Saier, *Microbiol. Rev.*, 1993, **57**, 320.
21. D. R. Liu and P. G. Schultz, *Proc. Natl. Acad. Sci. USA*, 1993, **96**, 4780.
22. H. Jakubowski and E. Goldman, *Microbiol. Rev.*, 1993, **56**, 412.
23. A. R. Fersht, *Biochemistry*, 1977, **16**, 1025.
24. A. R. Fersht and C. Dingwall, *Biochemistry*, 1979, **18**, 2627.

25. O. Nureki, D. Vassylyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M.Konno, T. Hendrickson, P. Schimmel and S. Yokoyama, *Science*, 1998, **280**, 578.
26. L. Silvian, J. Wang and T. Steitz, *Science*, 1999, **285**, 1074.
27. J. M. Pezzuto and S. M. Hecht, *J. Biol. Chem.*, 1980, **255**, 865.
28. G. Baldini, B. Martoglio, A. Schachenmann, C. Zugliani and J. Brunner, *Biochemistry*, 1988, **27**, 7951.
29. J. D. Bain, C. G. Glabe, T. A. Dix, and A. R. Chamberlin, *J. Am. Chem. Soc.*, 1989, **111**, 8013.
30. J. Ellman, D. Mendel, S. Anthony-Cahill, C. J. Noren and P. G. Schultz, *Methods Enzymol.*, 1991, **202**, 301.
31. R. Thompson, *TIBS*, 1988, **13**, 91.